

日 本 国 特 許 庁
JAPAN PATENT OFFICE

16.11.2004

RECD 13 JAN 2005

WIPO

PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2003年11月21日

出 願 番 号
Application Number: 特願2003-392845
[ST. 10/C]: [JP2003-392845]

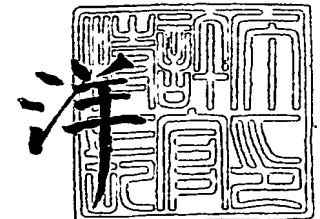
出 願 人
Applicant(s): 株式会社テクノネットワーク四国
株式会社エイアンドティー

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2004年12月24日

特許庁長官
Commissioner,
Japan Patent Office

小 川



BEST AVAILABLE COPY

【書類名】 特許願
【整理番号】 PA-15861
【提出日】 平成15年11月21日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 17/30
【発明者】
 【住所又は居所】 高知県南国市岡豊町小蓮 高知大学医学部付属病院内
 【氏名】 片岡 浩巳
【特許出願人】
 【識別番号】 803000012
 【氏名又は名称】 株式会社テクノネットワーク四国
【代理人】
 【識別番号】 100089118
 【弁理士】
 【氏名又は名称】 酒井 宏明
【手数料の表示】
 【予納台帳番号】 036711
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索装置であって、

前記白血球粒度パターンは、複数の細胞成分分画を含み、

実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段、

前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定手段、

前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定手段、

前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング手段、

前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成手段、

前記2次クラスマップと前記クラス間距離マスターを記憶する記憶手段、

被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定手段、および

前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索手段、

を備えたことを特徴とする類似パターン検索装置。

【請求項 2】

複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法であって、

前記白血球粒度パターンは、複数の細胞成分分画を含み、

実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリングステップ、

前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定ステップ、

前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定ステップ、

前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリングステップ、

前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成ステップ、

前記2次クラスマップと前記クラス間距離マスターを記憶する記憶ステップ、

被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記 2 次クラスマップの中から決定するクラス決定ステップ、および

前記クラス決定ステップで決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索ステップ、

を備えたことを特徴とする類似パターン検索方法。

【請求項 3】

複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法をコンピュータに実行させるプログラムであって、

前記白血球粒度パターンは、複数の細胞成分分画を含み、

実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1 次クラスマップを作成する 1 次クラスタリングステップ、

前記 1 次クラスマップに含まれる各パターンについて、所定の初期値を用いて EM アルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第 1 の混合分布モデルパラメータを決定する第 1 のパラメータ決定ステップ、

前記各白血球粒度パターンについて、前記第 1 の混合分布モデルパラメータを初期値として EM アルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第 2 の混合分布モデルパラメータを決定する第 2 のパラメータ決定ステップ、

前記各白血球粒度パターンについて、前記第 2 の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2 次クラスマップを作成する 2 次クラスタリングステップ、

前記 2 次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成ステップ、

前記 2 次クラスマップと前記クラス間距離マスターを記憶する記憶ステップ、

被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記 2 次クラスマップの中から決定するクラス決定ステップ、および

前記クラス決定ステップで決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索ステップ、

をコンピュータに実行させることを特徴とする類似パターン検索プログラム。

【請求項 4】

複数の細胞成分分画を含む白血球粒度パターンにおいて各細胞成分分画を分離する細胞成分分画分離装置であって、

実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1 次クラスマップを作成する 1 次クラスタリング手段、

前記 1 次クラスマップに含まれる各パターンについて、所定の初期値を用いて EM アルゴリズムを実行することにより、各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分分画の平均値、分散および密度からなる混合分布モデルパラメータを決定するパラメータ決定手段、および

前記各白血球粒度パターンについて、前記混合分布モデルパラメータを初期値として EM アルゴリズムを実行することにより、各白血球粒度パターンに含まれる各細胞成分の分画を分離する分画分離手段

を備えたことを特徴とする分画分離装置。

【書類名】明細書

【発明の名称】白血球粒度パターンの類似パターン検索装置、類似パターン検索方法、および類似パターン検索プログラム、および、分画分離装置

【技術分野】**【0001】**

本発明は、複数の細胞成分分画を含む白血球粒度パターンについて、複数の白血球粒度パターンを含む集団の中から、被検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索装置、および、白血球粒度パターンにおいて各細胞成分分画を分離する分画分離装置に関する。

【背景技術】**【0002】**

フローサイトメトリーは白血球を好中球、リンパ球、単球、好酸球などに短時間で分類することが可能な検査である。フローサイトメトリーによって得られる白血球粒度データは、細胞の成熟度や疾患によってさまざまな粒度パターンに分類することができる（非特許文献1参照）。

【0003】

この検査は、日常のスクリーニング検査法として多くの施設で導入されているが、分類された数値データのみが利用され、分析装置内部で発生する白血球粒度データを臨床で診断に利用することはまれであった。これは、白血球粒度データが膨大なデータ量であり、外部の情報システムで取り扱うことができなかった点と、分析生データに対して視覚的な探索方法しか実施されておらず、科学的な手法で検討することが困難であったためである。

【0004】

これに対し、本発明者らは、2次元ヒストグラムとして得られる白血球粒度データを用いて、自己組織化マップ(SOM)によりクラスタリングを行う方法を開発した（非特許文献2～4参照）。当該分類方法は、白血球粒度データをデータベースに記録し、データマイニングを適用して特徴的なパターンを抽出することにより、2次元ヒストグラムだけの情報では判断できなかった分類が可能となる。

【0005】

従来の分類方法は、分析装置内部で各分画の谷間を境界とした分離方法で処理されており、各分画を1つの数値データとして診断に利用する方法が用いられていた。しかし、この方法では、複数のクラスターが近接した分布、たとえば、好中球に属する桿状核球と分葉核球、あるいは、正常細胞と幼稚球の分離ができないという問題があった。

【0006】

【非特許文献1】 巽典之、津田泉、田窪考行、他：自動白血球分類結果の实地診療への反映、HORIBA Technical Reports、No.20、pp.23-26、2000。

【非特許文献2】 片岡浩巳、井沖浩美、小西修、他：白血球粒度のデータマイニング支援システムの構築、日本臨床検査自動化学会誌、Vol27、4、pp.583、2002。

【非特許文献3】 片岡浩巳、井沖浩美、小西修、他：白血球粒度のクラスタリングと3Dビジュアライゼーション、医療情報学 22 (Suppl.)、pp.209-210、2002。

【非特許文献4】 井沖浩美、片岡浩巳、川崎由夏、他：白血球粒度データによるアレルギー疾患領域のパターン分類、医療情報学 22 (Suppl.)、pp.211-212、2002。

【発明の開示】**【発明が解決しようとする課題】****【0007】**

本発明は、上記に鑑みてなされたものであって、白血球の細胞成分ごと、あるいは、各成分の組み合わせという総合的な視点で高速な類似検索を可能とするアルゴリズムを開発し、診断に有用な情報を提供する方法を提供することである。

【課題を解決するための手段】**【0008】**

本発明者らは、上記課題に鑑み、鋭意研究の結果、白血球粒度パターンに含まれる各細胞成分について、EMアルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能にできることを見だし、この知見に基づいて本発明を完成させるに至った。

【0009】

請求項1にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索装置であって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定手段、前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定手段、前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリング手段、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成手段、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶手段、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定手段、および、前記クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索手段、を備えたことを特徴とする。

この請求項1にかかる発明によれば、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

【0010】

請求項2にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法であって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリングステップ、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定ステップ、前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定ステップ、前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリングステップ、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを

作成するクラス間距離マスター作成ステップ、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶ステップ、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定ステップ、および、前記クラス決定ステップで決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索ステップ、を備えたことを特徴とする。

この請求項2にかかる発明によれば、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

【0011】

請求項3にかかる発明は、複数の白血球粒度パターンを含む集団の中から、被検検体の白血球粒度パターンと類似性の高いパターンを持つ白血球粒度パターンを検索する類似パターン検索方法をコンピュータに実行させるプログラムであって、前記白血球粒度パターンは、複数の細胞成分分画を含み、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリングステップ、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定する第1のパラメータ決定ステップ、前記各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定する第2のパラメータ決定ステップ、前記各白血球粒度パターンについて、前記第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成する2次クラスタリングステップ、前記2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、各クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するクラス間距離マスター作成ステップ、前記2次クラスマップと前記クラス間距離マスターを記憶する記憶ステップ、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するクラス決定ステップ、および、前記クラス決定ステップで決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類似性の高いパターンとして判定する類似パターン検索ステップ、をコンピュータに実行させることを特徴とする。

この請求項3にかかる発明によれば、自己組織化マップを適用して決定した初期値を用いてEMアルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2次クラスマップおよびクラス間距離マスターを構築する。

【0012】

請求項4にかかる発明は、複数の細胞成分分画を含む白血球粒度パターンにおいて各細胞成分分画を分離する細胞成分分画分離装置であって、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成する1次クラスタリング手段、前記1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分分画の平均値、分散および密度からなる混合分布モデルパラメータを決定するパラメータ決定手段、および前記各白血球粒度パターンについて、前記混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、各白血球粒度パターンに含まれる各細胞成分の分画を分離する分画分離手段を

備えたことを特徴とする。

この請求項 4 にかかる発明によれば、EM アルゴリズムの初期値の決定に自己組織化マップ(SOM)を適用する。

【発明の効果】

【0013】

本発明(請求項 1)にかかる類似パターン検索装置は、自己組織化マップを適用して決定した初期値を用いて EM アルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2 次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

【0014】

従来は、2 次元ヒストグラムの粒度データを直接用いて SOM によりクラスタリングを行っていたため、白血球の個々の成分に注目した部分的な類似性に注目した類似検索を行うことができなかった。本発明によれば、EM アルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能とするものである。

【0015】

また、本発明(請求項 2)にかかる類似パターン検索方法は、自己組織化マップを適用して決定した初期値を用いて EM アルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2 次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

【0016】

また、本発明(請求項 3)にかかる類似パターン検索プログラムは、自己組織化マップを適用して決定した初期値を用いて EM アルゴリズムにより白血球粒度の各成分を分離し、自己組織化マップを用いて再びクラスタリングを行うことにより、2 次クラスマップおよびクラス間距離マスターを構築するため、検索対象の類似性を自由に選択できるという効果を奏する。

【0017】

また、本発明(請求項 4)にかかる分画分離装置は、EM アルゴリズムの初期値の決定に自己組織化マップ(SOM)を適用するため、周辺尤度の局所的最大値への収束問題を解決できるという効果を奏する。

【0018】

一般的に EM アルゴリズムは、収束点が初期条件に強く依存し、周辺尤度の局所的最大値を回避できない場合が存在する問題を抱えている。つまり、初期値によっては、低品質の局所解に収束するという現象がある。本発明では、この問題に対し、あらかじめ全体の白血球粒度データを SOM によりクラスタリングした結果をもとに、各クラスの初期値を求め周辺尤度の局所的最大値への収束問題を解決するものである。

【発明を実施するための最良の形態】

【0019】

以下に、本発明の実施形態について説明する。図 1 は、本実施の形態にかかる類似パターン検索装置 1 の構成を示すブロック図である。本実施の形態にかかる類似パターン検索装置 1 は、1 次クラスタリング部 11、第 1 のパラメータ決定部 12 第 2 のパラメータ決定部 13、2 次クラスタリング部 14、クラス間距離マスター作成部 15、メモリ 16、クラス決定部 17、および類似パターン検索部 18 を備えている。

【0020】

本発明は、EM アルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能とすることを特徴とする。

【0021】

EMアルゴリズムはExpectation step(E-step)とMaximization step(M-step)の2つの処理アルゴリズムから構成され、これらの操作を収束するまで繰り返してパラメータを更新することによって、最尤推定量の極大点を得ることができる。E-stepは、対数尤度の条件付期待値を計算し、M-stepは、条件付期待値を最大化する処理が行われる。

本実施の形態で用いたデータセットと近似モデルは、

データタイプ: 2次元ヒストグラム

モデル: 正規混合モデル

パラメータ: 平均、分散、密度
である。

【0022】

一般的にEMアルゴリズムは、収束点が初期条件に強く依存し、周辺尤度の局所的最大値を回避できない場合が存在する問題を抱えている。つまり、初期値によっては、低品質の局所解に収束するという現象がある。本発明では、この問題に対し、あらかじめ全体の白血球粒度データをSOMによりクラスタリングした結果をもとに、各クラスの初期値を求め周辺尤度の局所的最大値への収束問題を解決するものである。

【0023】

分析装置2で測定した白血球粒度の2次元ヒストグラムデータは、類似パターン検索装置1に送信され、メモリ16に格納される。

【0024】

1次クラスタリング部11は、実測して得られた複数の前記白血球粒度パターンに自己組織化マップを適用してクラスタリングを行い、1次クラスマップを作成するものである。

【0025】

第1のパラメータ決定部12は、1次クラスマップに含まれる各パターンについて、所定の初期値を用いてEMアルゴリズムを実行することにより、前記各パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第1の混合分布モデルパラメータを決定するものである。

【0026】

第2のパラメータ決定部13は、実測した各白血球粒度パターンについて、前記第1の混合分布モデルパラメータを初期値としてEMアルゴリズムを実行することにより、前記各白血球粒度パターンに含まれる細胞成分の成分数、ならびに、各細胞成分の平均値、分散および密度からなる第2の混合分布モデルパラメータを決定するものである。

【0027】

2次クラスタリング部14は、第2の混合分布モデルパラメータに自己組織化マップを適用してクラスタリングを行い、2次クラスマップを作成するものである。

【0028】

クラス間距離マスター作成部15は、2次クラスマップに含まれる各クラス間におけるすべての組合せの類似距離を計算し、クラスの組合せと当該クラス間の類似距離とを対応させたクラス間距離マスターを作成するものである。

【0029】

メモリ16は、分析装置2で測定した白血球粒度の2次元ヒストグラムデータ、2次クラスタリング部14で作成した2次クラスマップデータ、クラス間距離マスター作成部15で作成したクラス間距離マスターデータ等のデータを記憶するものである。

【0030】

クラス決定部17は、被検検体の白血球粒度パターンに含まれる各細胞成分分画に属するクラスを前記2次クラスマップの中から決定するものである。

【0031】

類似パターン検索部18は、クラス決定工程で決定されたクラスとの類似距離が所定の閾値以下であるクラスを前記クラス間距離マスターの中から類似クラスとして検出し、当該類似クラスに含まれる白血球粒度パターンを、前記被検検体の白血球粒度パターンと類

似性の高いパターンとして判定するものである。

【0032】

外部入出力装置2は、ユーザによって入力された各種パラメータや類似パターンの検索条件等を類似パターン検索装置1に送信する。また、類似パターン検索装置1でヒットした類似パターンを画面上に出力する。

【0033】

図2に本実施の形態にかかる類似パターン検索装置1によって行われる処理のフローチャートを示す。なお、以下の説明では、処理の具体例として、自動血球計数装置 PENTRA120(堀場製作所)により分析された一般患者検体8,800件のLMNEチャンネルの2次元ヒストグラム128*128、8bit/検体のデータを処理するケースに沿って説明を行う。

【0034】

(1) 初期値決定のための1次マップの作成

分析装置2から出力された2次元ヒストグラムデータを近傍8点の平滑化処理を行ったデータに関して、入力層128*128(16,384ニューロン)、競合層12*12(ユニット)のSOMでクラスタリングを行い、得られた144種類のパターンを1次クラスマップとした。SOMの学習パラメータは、近傍距離4、学習率0.3とした。さらに、この1次クラスマップの個々のパターンについて、4*4の16分割した領域を設定し、それぞれの2次元ヒストグラムの重心を求め、その重心を初期値として、EMアルゴリズムにより混合モデルの分離を行った。各分画の分布モデルは正規分布と仮定して計算を行った。さらに、得られた混合分布モデルパラメータ(成分数、各成分の平均値、分散、密度)を人為的に調整し一時的なパラメータを決定した。

【0035】

(2) EMアルゴリズムによる混合分布近似

EMアルゴリズムによる混合分布近似については、「データ学習アルゴリズム」(渡辺澄夫、共立出版、2001.)およびIgor V. Cadez, Scott Gaffney, Padhraic Smyth: A General Probabilistic Framework for Clustering Individuals and Objects, Knowledge Discovery and Data Mining, pp140-149, 2000.等に記載の手法を用いて実施することができる。

【0036】

具体的には、各検体の2次元ヒストグラムデータに対して、1次マップに最も類似したクラスを検索し、そのクラスの混合分布モデルパラメータのパラメータを使ってEMアルゴリズムを実行し粒度成分を分離した。全ての被検粒度データについて同じ処理を実行し個別の混合分布モデルパラメータを計算した。

【0037】

(3) 混合分布パラメータを入力としたSOMによる2次マップの作成

混合分布パラメータを入力としたSOMによる2次マップの作成については、Tom Heskes: Self-organizing maps, vector quantization, and mixture modeling, IEEE Transactions on Neural Networks, 12: pp1299-1305, 2001.等に記載の手法を用いて実施することができる。

【0038】

競合層30*30(ユニット)、近傍距離10、学習率0.3のSOMを用いて、X平均値、Y平均値、X共分散行列、Y共分散行列、XY共分散行列、そして、密度の6個で構成された実数の混合分布モデルパラメータを入力層に加えクラスタリングを行った。このクラスタリング結果を2次マップとして類似検索に用いた。この時、各クラス間における全ての組み合わせの類似距離を求めておき、クラス間距離マスターに登録した。

【0039】

(4) 類似検索

被検検体の各分画に属するクラスを2次マップから求め、クラス間距離マスターを読み込み、検索の目的にあわせて閾値を決定し、その条件に一致するクラス群を検索した。閾値を可変することにより、検索の類似性の強度を自由に選択できるようにし、さらに、閾

値に含まれる領域のクラス群について選言条件で検索することにより類似検索を実現した。各分画の総合的なパターンについての検索を行いたい場合は、それぞれの分画に属するクラスの連言により検索することとした。

【0040】

図3にSOMにより1次クラスタリングを行った結果を示す。12*12の競合層の内部を表示したもので、白血球粒度の全体のパターンを144個にクラスタリングされた結果が得られた。

【0041】

図4の上図は、もとの粒度データの2次元ヒストグラムを示し、+が初期値、×がEMアルゴリズムにより最適尤度の探索が行われた経路と収束点を示す。図4の下図は、得られた混合分布パラメータを用いて各分画成分を合成して再描画を行ったモデル化された2次元ヒストグラムである。

【0042】

図5は、EMアルゴリズムにより得られた個々の混合分布モデルパラメータをSOMでクラスタリングした結果を示す。赤で描画した楕円形の成分は、1成分の細胞の分画を示し、周囲に類似したパターンが配置された結果が得られた。それぞれの細胞群についてさまざまなパターンが存在していることが理解できる。ピンク色1はリンパ球、黄色2は単球、水色3で示す領域は好中球、紫色4は好酸球の分布結果が得られた。文字どおりのLMNEチャンネルの4つの細胞群がクラスタリングされた結果を得た。さらに、リンパ球の下部に分布した白色領域は血小板がマッピングされており、その他の白色部分と各細胞群の境界領域には異常細胞と考えられる分布がマッピングされた。図5と図6に示した細胞群は、左上の角をClass0とし、右下の角をClass899としたラスタ方向のシーケンシャル番号で呼ぶことにした。

【0043】

図6は、好中球領域に分布する桿状核球と分葉核球の分布を示す。顕微鏡による目視分類結果よりClass120は桿状核球がどのクラスよりも多いクラスで、Class351は、分葉核球が多い症例のクラスである。黄色31（左）のグラディエーション領域は、核左方移動が著明な桿状核球が一番多く含まれる症例群であったClass120を中心とした類似距離のパターンを色の強度で表現した分布である。また、青色32（右）のグラディエーション領域は、分葉核球が一番多かったClass351を中心としたパターンを示している。

【0044】

図7は、Class351を基準とした分葉核球の分布を拡大した図である。幅広い範囲で類似検索を行いたい場合、赤線で囲んだ領域のクラスを対象に検索し、類似性が強い細胞の検索を行いたい場合は、緑線、あるいは、青線で囲まれた領域のクラスを検索することで検索対象の絞り込みが可能となった。

【0045】

図6で示した桿状核球(Class120)と分葉核球(Class351)との視覚的な境界面は、滑らかなグラディエーションで結合されており、類似性の境界が不明瞭であることを示している。これは、桿状核球と分葉核球が同じ好中球に属することから、細胞の分化度の視点から見た類似性がマップ上でクラスタリングができていていることを示唆している。一方、Class351の分葉核球とリンパ球領域との境界はグラディエーションの少ない明瞭な境界面が観察され、これらの細胞群はマップ上で明確に分離できることを示している。図8は、好酸球のClass801を基準とした各クラスの距離をプロットした図である。縦軸は、Class801からの距離、横軸は、距離の少ない順にソートしたクラスである。距離が1以下は同じ好酸球が分布しており、距離の閾値を可変することで、検索対象の類似性を可変できることを示している。また、細胞ごとに階段状の曲線が得られ、好中球の分葉核球と桿状核球が単球で分断されている興味深い結果が得られた。これは、基準となる細胞によってさまざまなパターンとなる傾向があった。

【0046】

白血球の各成分単独、あるいは、各成分を統合した類似性に関して、類似性の尺度を自

由に可変できる類似検索システムを構築した。EMアルゴリズムは、事前にSOMによりクラスタリングしたパターンで初期値を決定することにより、正しい収束結果が得られた。また、臨床検査領域のフローサイトメトリーで桿状核球と分葉核球の分離は不可能であったが、本法を用いることで容易に分離することが可能となり、診断や治療に有用な情報を提供するシステムを構築することができた。

【産業上の利用可能性】

【0047】

以上のように、本発明にかかる類似パターン検索装置は、各成分を統合した類似性に関して、類似性の尺度を自由に可変できるため、診断や治療に有用な情報を提供することができる。

【図面の簡単な説明】

【0048】

【図1】本実施の形態にかかる類似パターン検索装置1の構成を示すブロック図である。

【図2】本実施の形態にかかる類似パターン検索装置1によって行われる処理のフローチャートである。

【図3】SOMにより1次クラスタリングを行った結果得られた1次クラスマップの一例を示す図である。

【図4】もとの粒度データの2次元ヒストグラム（上図）、および、得られた混合分布パラメータを用いて各分画成分を合成して再描画を行ったモデル化された2次元ヒストグラム（下図）である。

【図5】EMアルゴリズムにより得られた個々の混合分布モデルパラメータをSOMでクラスタリングした結果得られた2次クラスマップの一例を示す図である。

【図6】好中球領域に分布する桿状核球と分葉核球の分布を示す図である。

【図7】Class351を基準とした分葉核球の分布の拡大図である。

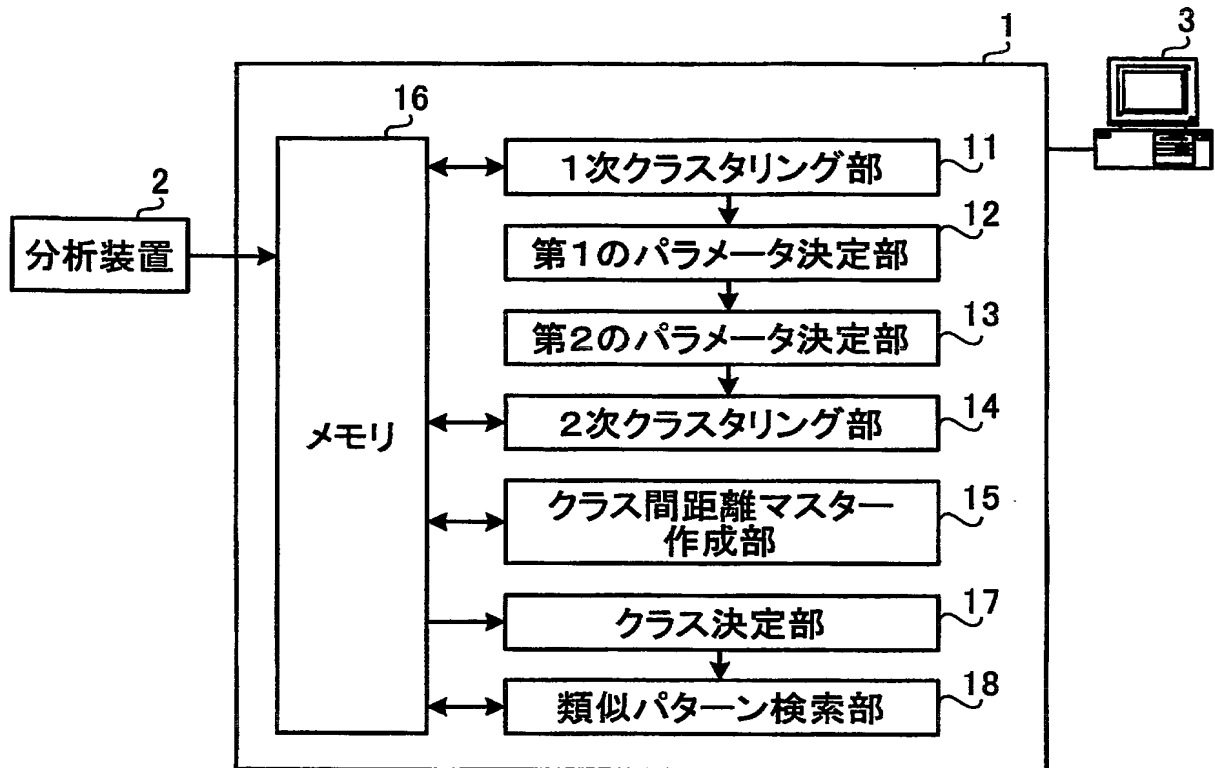
【図8】好酸球のClass801を基準とした各クラスの距離をプロットした図である。

【符号の説明】

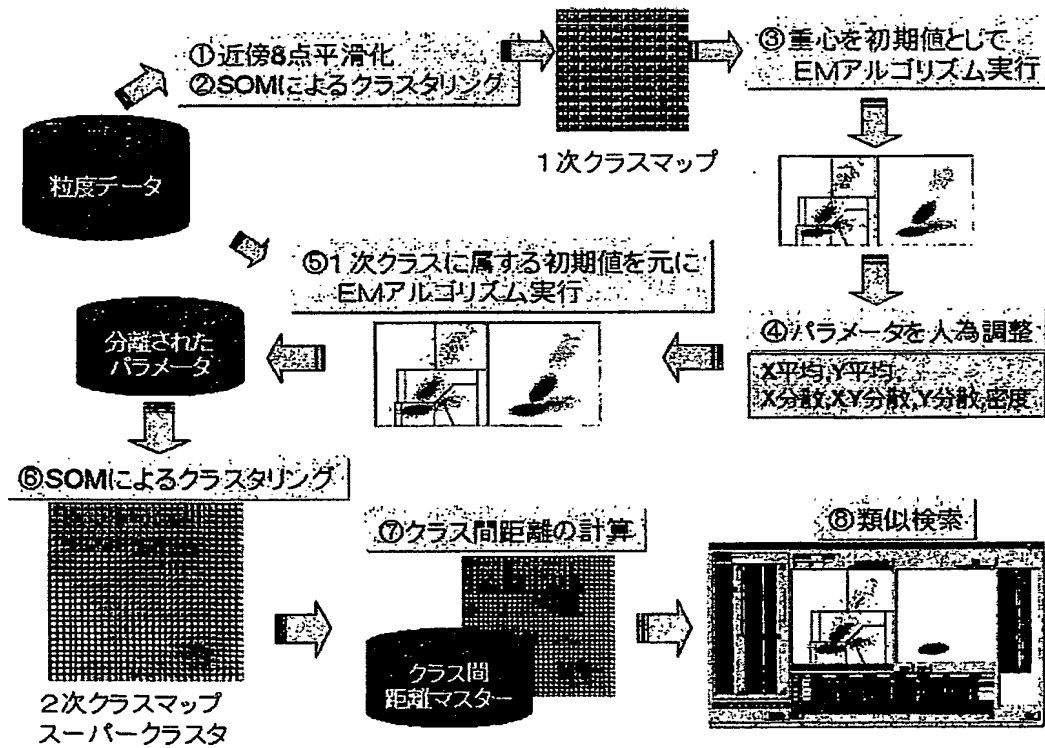
【0049】

- 1 類似パターン検索装置
 - 11 1次クラスタリング部
 - 12 第1のパラメータ決定部
 - 13 第2のパラメータ決定部
 - 14 2次クラスタリング部
 - 15 クラス間距離マスター作成部
 - 16 メモリ
 - 17 クラス決定部
 - 18 類似パターン検索部
- 2 分析装置
- 3 外部入出力装置

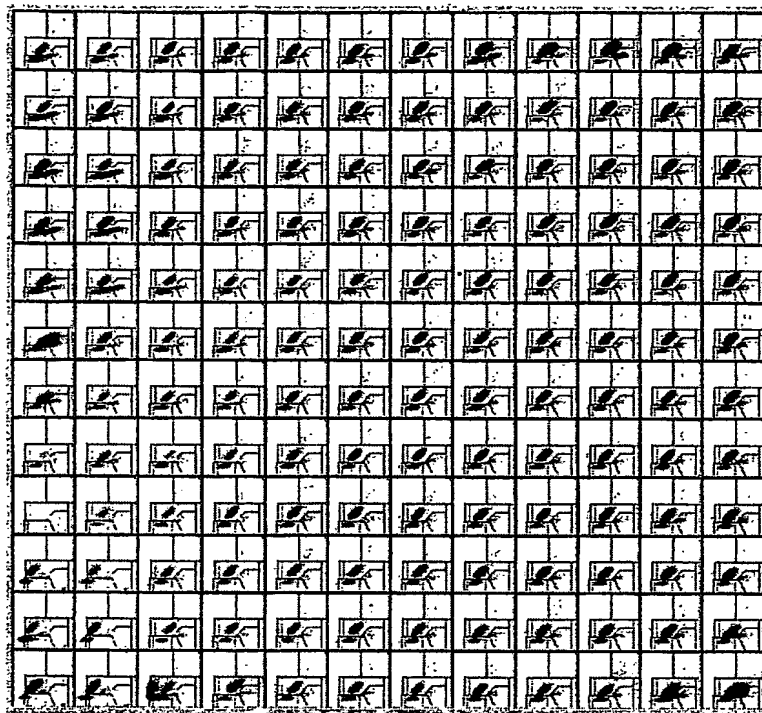
【書類名】 図面
【図 1】



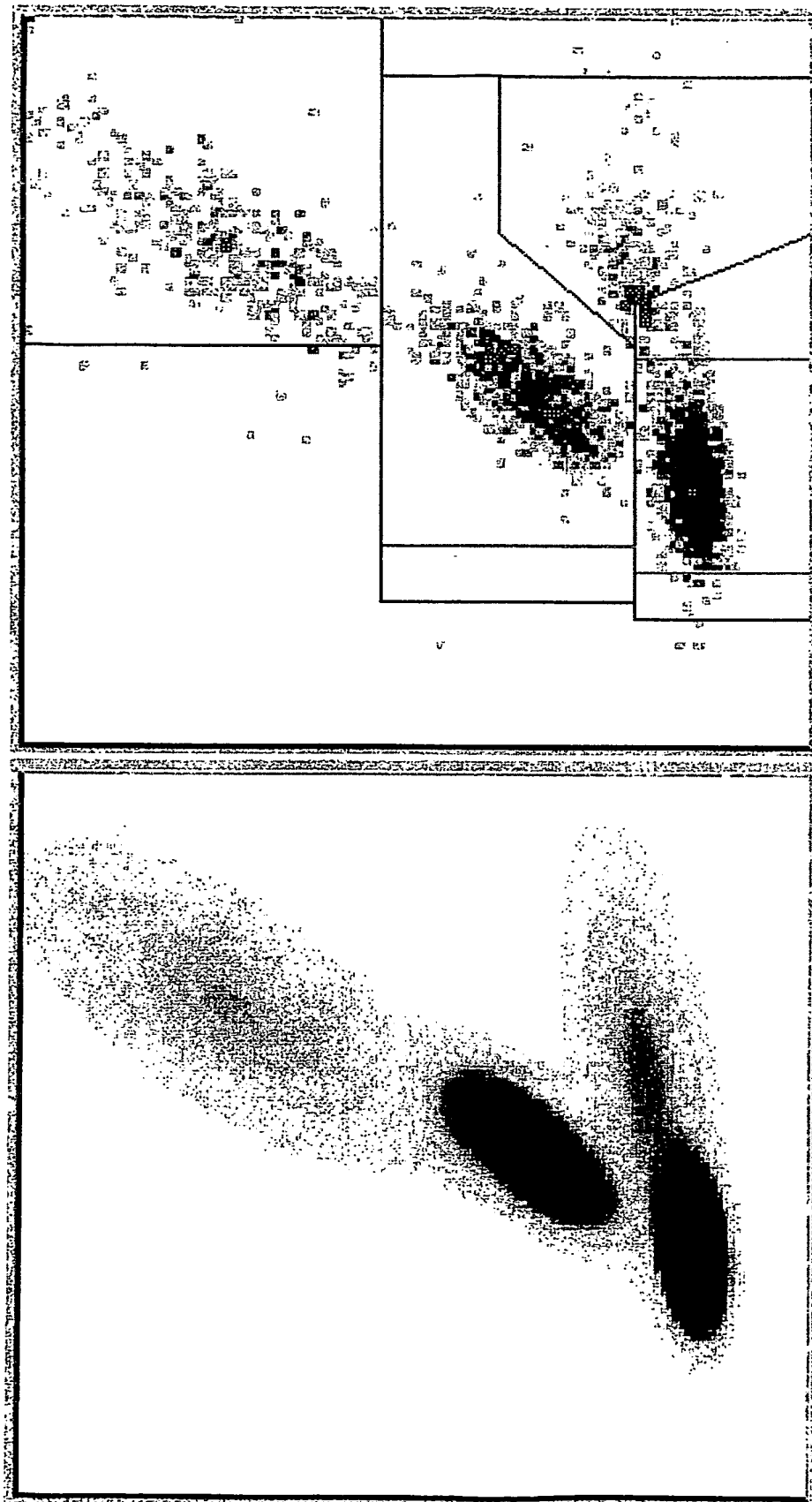
【図 2】



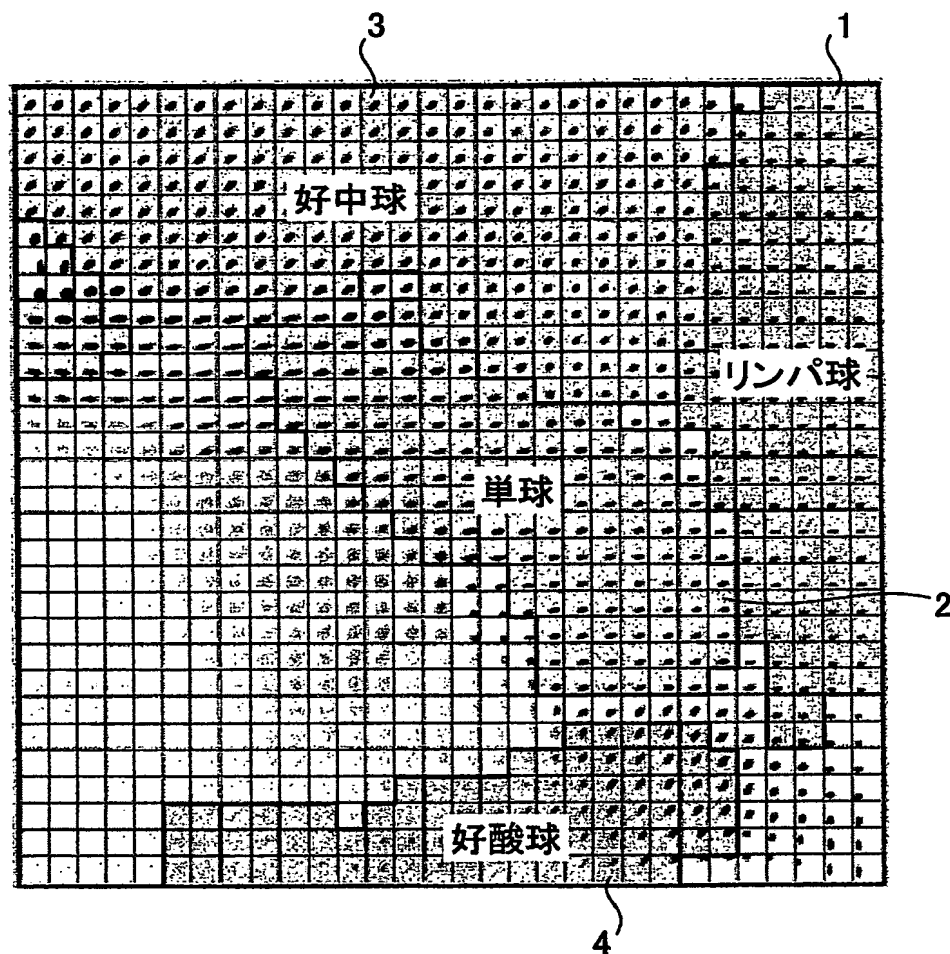
【図 3】



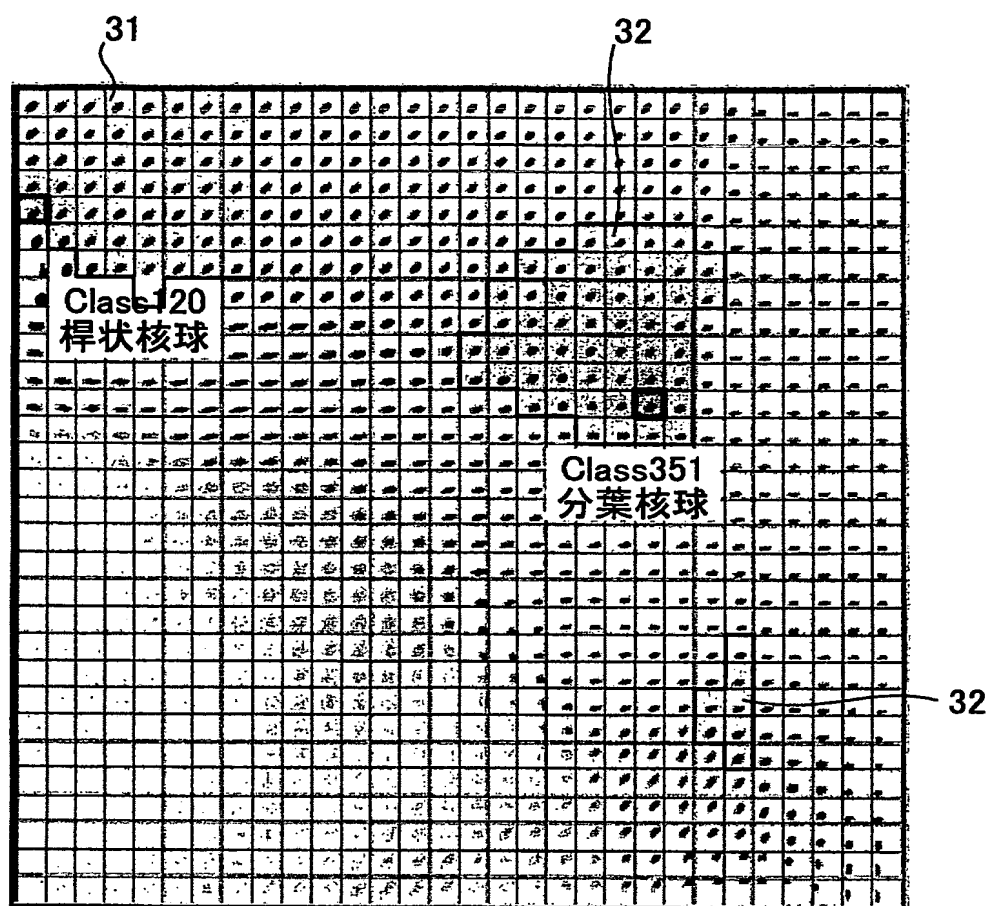
【图 4】



【図 5】

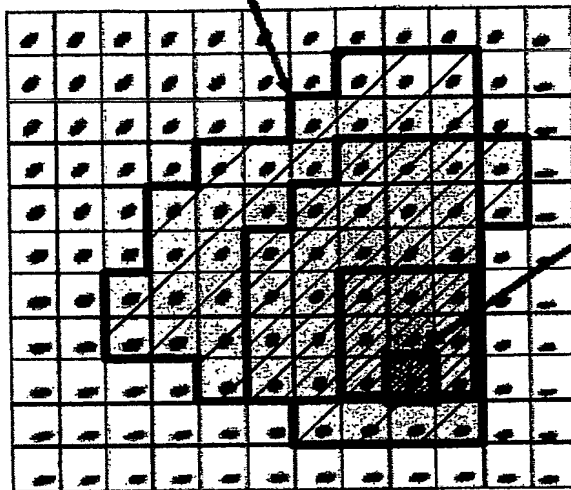


【図 6】



【図7】

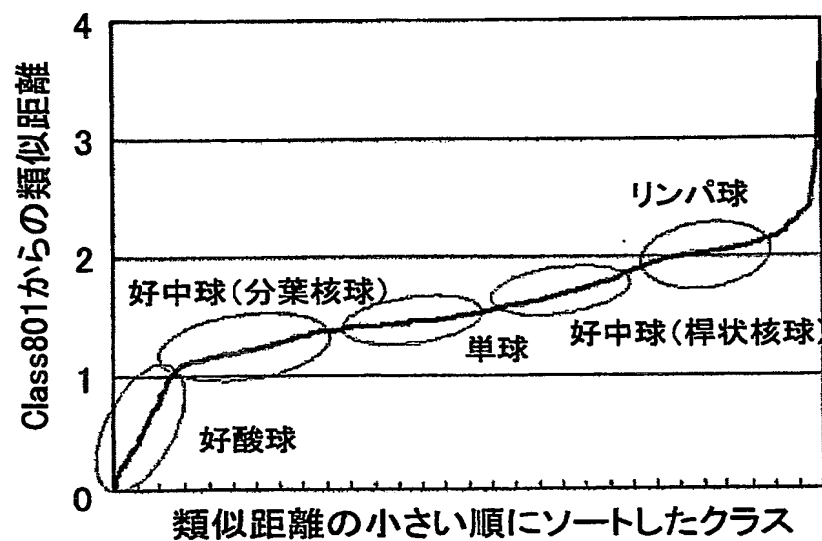
Class351を基準とした広い類似距離範囲



Class351を基準とした
狭い類似距離範囲

※各クラスの選言により検索の類似性を可変

【図8】



【書類名】 要約書

【要約】

【課題】 白血球の細胞成分ごと、あるいは、各細胞成分の組み合わせという総合的な視点で高速な類似検索を可能とするアルゴリズムを開発し、診断に有用な情報を提供する。

【解決手段】 複数の細胞成分分画を含む白血球粒度パターンにおいて、EMアルゴリズムを用いた混合密度近似を行うことで各成分を分離し、さらに、各分画の特徴パラメータをクラスタリングすることで、目的の細胞群の分布パターンに注目した類似検索を可能とする。

【選択図】 図 2

【書類名】 出願人名義変更届
【整理番号】 PA-15861
【提出日】 平成16年10月20日
【あて先】 特許庁長官 殿
【事件の表示】
 【出願番号】 特願2003-392845
【承継人】
 【識別番号】 591258484
 【氏名又は名称】 株式会社エイアンドティー
【承継人代理人】
 【識別番号】 100089118
 【弁理士】
 【氏名又は名称】 酒井 宏明
【手数料の表示】
 【予納台帳番号】 036711
 【納付金額】 4,200円
【提出物件の目録】
 【包括委任状番号】 0001282
 【物件名】 譲渡証書 1
 【援用の表示】 手続補足書にて提出の譲渡証書

認定・付加情報

特許出願の番号	特願 2003-392845
受付番号	50401789278
書類名	出願人名義変更届
担当官	鈴木 康子 9584
作成日	平成 16 年 11 月 29 日

< 認定情報・付加情報 >

【承継人】

【識別番号】 591258484

【住所又は居所】 神奈川県藤沢市遠藤 2023 番地 1

【氏名又は名称】 株式会社エイアンドティー

【承継人代理人】 申請人

【識別番号】 100089118

【住所又は居所】 東京都千代田区霞が関 3 丁目 2 番 6 号 東京倶楽
部ビルディング 酒井国際特許事務所

【氏名又は名称】 酒井 宏明

特願 2003-392845

出 願 人 履 歴 情 報

識別番号 [803000012]

1. 変更年月日	2002年12月26日
[変更理由]	新規登録
住 所	香川県高松市丸の内2番5号
氏 名	株式会社テクノネットワーク四国

特願 2 0 0 3 - 3 9 2 8 4 5

出 願 人 履 歴 情 報

識別番号 [5 9 1 2 5 8 4 8 4]

1. 変更年月日 2 0 0 1 年 1 0 月 1 9 日

[変更理由] 住所変更

住 所 神奈川県藤沢市遠藤 2 0 2 3 番地 1

氏 名 株式会社エイアンドティー

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.